How to choose and evaluate a point-of-care testing instrument

Introduction

Point-of-care testing (PoCT) instruments come in variety of shapes and sizes and range from simple testing strips such as those used for glucose measurement to small desk-top devices that can accurately measure tests such as glycated haemoglobin (HbA1c). As devices become larger and complex, more consideration and thought is required to determine which is the best device for your need and the environment in which you work.

Selecting and possibly evaluating PoCT devices can be considered in two phases. The first is a general consideration of the features and performance of a particular device as stated by the supplier of the device, and whether these features meet your particular needs.

The second phase is an assessment of the analytical capability of the device using well-established methods including statistical analysis of the results. Assessment of analytical capability or the ability of the device to measure precisely and accurately is a process that is performed regularly as part of routine PoCT. This is because, as stated in the NPAAC POCT standards: "*The quality of PoCT may be affected by many factors, including the storage of consumables, PoCT practitioners, specimen quality and variability between instruments. Your practice needs to assess the impact of these factors by evaluating the analytical performance of your PoCT system"* (1).

This regular evaluation process is called quality control or quality management and information about these procedures can be found on the APPN website and are documented in the NPAAC and RACGP PoCT Standards (1,2).

The full analytical evaluation described later goes beyond these routine quality management procedures and includes assessment of other factors which affect test results such as whether the analytical method is linear or gives a consistent response over the expected testing range, and whether the method is subject to interference or falsely positive or negative results due to the presence in the blood of particular substances.

It is unlikely that an individual PoCT provider such as a general practice will have the resources, or indeed the need, to carry out a complete analytical evaluation although an individual practice or practices may participate in parts of an evaluation organised by another organisation such as APPN.

Furthermore unless the device is new and making its appearance in the market for the first time there will be information about the analytical performance of instruments available from APPN such as through its HbA1C device register or in the peer-reviewed literature. A good example of such an evaluation is described by Lenters-Westra et al. (3).

Important specifications to be considered when choosing a PoCT device

The device supplier or manufacturer will provide a product description including what are known as its specifications. The features and specifications that need to be considered include all of the following; some of them will not be relevant to smaller strip type devices:

- The test or tests measured
- The particular population group for which the device is intended
- The analytical principles of each test
- The weight and overall dimensions
- The power requirements
- Whether refrigeration of the test reagents or consumables are required
- The sample type e.g. whole blood, plasma, other fluid
- The volume of sample required
- Whether it is a qualitative (yes or no) or quantitative (a number or concentration is provided) test
- How it is calibrated and whether traceability of the calibration is provided
- The measurement range e.g. for HbA1c it might be 20-100 mmol/mol
- The recommended operating temperature range
- The time taken to provide a result from turning on the instrument to getting a result
- The actual analysis time from placing a sample into the device to when a result is provided

- Quality controls that are required to be analysed and how often
- The precision, accuracy or sensitivity and specificity of the test
- Any substances that may interfere, or falsely increase or decrease the result
- Whether and how the instrument can be linked to a desktop computer or other IT system
- How any infectious waste generated by the testing process is disposed
- The maintenance requirements.
- Any limitations stated by the manufacturer
- The regulatory status including ARTG and FDA approval, and CE marking

Additional information about devices to help you interpret these specifications can be obtained from other practices that may be using the same device, from the APPN which recommends the use of certain devices or from the peer-reviewed medical literature.

A full analytical evaluation process

1. Introduction

The purpose of an evaluation, known as a verification process, is to confirm (verify) that a particular PoCT system performs to and meets the manufacturer's stated specifications, and to validate the results against a known standard. The method evaluation, validation and verification process provides the evidence that a method is fit for purpose or that the particular requirements for a specific intended use are fulfilled.

The evaluation strategy and statistical analysis used will depend on whether the PoCT result is quantitative (where an absolute level or concentration of analyte is detected, e.g. for most clinical chemistry and haematology tests) or qualitative (where a result is reported as positive or negative, e.g. for most infectious disease and drugs of abuse tests). The key measures of analytical performance for most quantitative PoCT are accuracy (trueness of the result) and precision (reproducibility of the measurement). For qualitative PoCT, sensitivity (true positive rate) and specificity (true negative rate) are more commonly used to assess performance.

2. *Preliminary steps*

i. The desired analytical quality goals or the precision and accuracy required of the device should be clearly stated. These will in turn be based on the intended clinical use of the test.

ii. The scope or extent of the evaluation should be described.

iii. Some familiarity should be gained with the device and its operation before commencing the evaluation including the resolution of any problems.

iv. Ideally some of the evaluation will also involve those PoCT operators who will be using the device routinely in their workplace.

3. *Testing process*

i. Precision

Measurement of precision is usually expressed numerically as imprecision – standard deviation (SD) or coefficient of variation (CV). Measurements of precision should include both quality control material and patient samples.

Both intra assay imprecision (within run) and inter assay (between run) imprecision should be determined. Intra assay imprecision is usually determined using the same lot number of consumables and the same operator over a short period of time. A minimum of 20 replicates should be used.

Interassay imprecision is determined over an extended period of time and involves multiple operators and different consumable lot numbers.

Imprecision goals – maximum allowable SD and/or CV(%) at the clinically relevant analyte concentration should be determined. It is recommended to perform one run per day with 3 replicates at each of two concentrations daily for 5 days (4). With patient samples only within run imprecision is usually possible due to the nature of the fresh human material. Calculate the standard deviation for replicates and compare to manufacturers

claims. The imprecision, expressed as a coefficient of variation [CV%], is calculated using the formula:

CV% = (standard deviation [SD]/mean) x 100%.

As a general rule, the lower the imprecision, the better the reproducibility of the device. If the samples used in the patient comparison are tested in duplicate on the PoCT device, then the imprecision can be estimated as follows: take the differences between the duplicates, square each difference, add up all the squares, divide by twice the number of pairs of results (total number of results) and take the square root. It is recommended that imprecision studies are also performed using quality control material.

ii. Linearity

Linearity is used to establish the measuring interval that can be reported for the assay under evaluation. The measuring interval includes all test results between lower and upper limits which can be reported and used clinically. When evaluating a PoCT method, the linearity quoted by the manufacturer should be confirmed by running a minimum of 2 replicates at 5-7 concentrations over the claimed measuring interval (5). Possible matrix effects or the influence of substances contained or not contained in the material being used compared to the material intended to be analysed, must be excluded.

iii. Method Comparison

If the device is compared to an approved reference method the difference between the PoCT device and the comparative method measures the trueness of the PoCT device. If the comparative method is not a reference method (e.g. a standard laboratory method) trueness cannot be claimed. Instead the bias reported will be “bias to comparative method”. The appropriate way to determine which method delivers the correct result is to compare to a reference method.

At least 40 samples covering the clinically meaningful range should be included in study. Duplicates should be run for both PoCT device and comparative method. Samples should be run within a time span consistent with analyte stability. In general, the time span should not

exceed 2 hours for analysis by each method. The manufacturers sample stability recommendations must be followed.

While running comparisons all errors should be documented. Any data attached to documented errors should not be included in the final analysis. Any cause that requires rejection of data should be documented. Any discrepant results must be further investigated using a third (different) laboratory method, When the data is plotted as a scattergram and non-constant scatter is observed or suspected, then it is recommended that more than 40 comparison samples are used for analysis.

There are various ways to analyse the data including the following:

a. Construct a Bland-Altman difference plot.

This plot will calculate the average (mean) bias of the PoCT device relative to the comparative method and the limits of agreement (limits within which 95% of the differences fall). The closer the mean bias is to zero the closer the POCT device is to the laboratory method. If a bias exists this plot will highlight if the bias is constant across the concentration range or if it is proportional to the analyte concentration. The bias should be within published clinically allowable limits (6)

b. Perform a regression analysis.

The correlation coefficient (r) characterises the dispersion of results around the line of best fit. The closer that r is to 1, then the better the fit. Potential reasons for obtaining low r values include:

- Inadequate range of values
- Interferences
- Poor correlation between methods

The slope (proportional bias) indicates the lean of the line of best fit – the closer to 1 the better; the extent of the slope is often related to calibration differences between the methods. The Y-intercept is the point at which the line of best fit intersects the y axis (constant bias) – the closer to zero the better; this may also be related to calibration

A Passing Bablok regression is recommended since it allows you to compare two analytical methods to determine systematic error or bias. The advantage of using a Passing Bablok regression over least squares linear regression is that it allows measurement error in both X and Y variables. This means it does not assume that measurement error is normally distributed and therefore making it more robust against outliers.

c. Error Grid analysis

A useful tool in evaluations of glucose meters is an error grid analysis. The error grid is useful to quantify the clinical accuracy of a glucose meter compared to a reference value. It is a clinically oriented nonparametric approach to comparing blood glucose methods, based on three clinical assumptions: 1) glucose readings <3.9 mmol/L should be increased, 2) glucose readings >10 mmol/L should be decreased, and 3) acceptably accurate estimates are within 20% of the reference method or when both the estimates and reference blood glucose values are <3.9 mmol/L.

The grid breaks down a scatterplot of a reference glucose method and an evaluated glucose meter into five regions:

- Region A are those values within 20% of the reference method
- Region B contains points that are outside of 20% but would not lead to inappropriate treatment
- Region C are those points leading to unnecessary treatment
- Region D are those points indicating a potentially dangerous failure to detect hypoglycemia or hyperglycemia
- Region E is those points that would confuse treatment of hypoglycemia for hyperglycemia and vice-versa.

iv. Interference

Interference is an artefactual over or under reporting of a result due to the presence of a substance that reacts non-specifically with the measuring system. Substances to be tested are selected from the manufacturer's performance claims or published reports on interfering substances which affect analyte of interest.

Two aliquots of an identical patient sample are required to test interference. The interfering substance being checked is added to one sample, the other sample has added to it a solution that does not contain interfering substance. Both samples are analysed to see if there is any difference in values due to addition of interfering substance. Some key principles are that:

- Samples should be analysed in duplicate.
- The amount of interfering substance added needs to achieve values near the expected maximum concentration expected in the patient population. If interference is found at maximum concentration then lower concentrations of interfering substance should be tested to determine the level at which the interference first affects test results.
- It is recommended at least three analyte concentrations are tested for interference

A good practice is to test (as the minimum) the following common interferences:

- Bilirubin – test by adding standard bilirubin solution
- Haemolysis – test by mechanically haemolysing part of one of the paired samples by freezing and thawing and adding back to the original at predetermined concentrations of haemoglobin.
- Lipemia – test by adding a commercial fat emulsion or by analysing a lipemic patient sample before and after ultracentrifugation
- Exogenous analytes/drugs – test by adding analyte/drug of interest

4. Conclusions to be drawn from the analytical evaluation

Some of the key points to consider when reviewing the results of the evaluation are:

i. How do they compare with other published evaluations of the same equipment including concordance to manufacturer's claims?
ii. How do they compare with recommendations from professional societies?
iii. Whether it is possible to recommend the clinical use/population group (if any) for which the instrument is suitable.

References

1. NPAAC. Guidelines for Point of Care Testing (PoCT) [Internet]. Australian Government Department of Health; 2021 [cited 2019 Aug 4]. Available from: https://www.health.gov.au/internet/main/publishing.nsf/Content/35DE5FC4786CBB33CA257EEB007C7BF2/$File/Guidelines%20PoCT%201st%20Ed%202015.pdf

2. Royal Australasian College of General Practitioners. RACGP - Standards for point-of-care testing [Internet]. 2018 [cited 2022 Jan 18]. Available from: https://www.racgp.org.au/running-a-practice/practice-standards/standards-5th-edition/point-of-care-testing-1

3. Lenters-Westra E, English E. Are hemoglobin A1c point-of-care analyzers fit for purpose? The story continues. Clin Chem Lab Med. 2021 Mar 26;59(4):765–74.

4. Clinical and Laboratory Standards Institute. EP15-A2 User verification of performance for precision and trueness.

5. Clinical and Laboratory Standards Institute. EP06A Evaluation of the linearity of quantitative measurement procedures.

6. RCPA QAP Allowable limits of performance Available from: https://www.rcpaqap.com.au/docs/2014/chempath/ALP.pdf